

#2

OIPE

RAW SEQUENCE LISTING

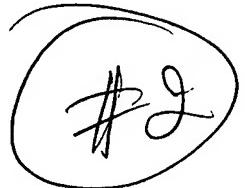
PATENT APPLICATION: US/10/003,759

DATE: 12/14/2001

TIME: 09:52:58

Input Set : A:\2739.2001-001 seq.txt
 Output Set: N:\CRF3\12142001\I003759.raw

ENTERED



3 <110> APPLICANT: Wicher, Kryzysztof B.
 4 Holst, Olof Peder
 5 Hachem, Maher Youssef Abou
 6 Karlsson, Eva Margareta Nordberg
 7 Hreggvidsson, Gudmundur O.
 9 <120> TITLE OF INVENTION: Thermostable Cellulase
 11 <130> FILE REFERENCE: P5099PC00
 C--> 13 <140> CURRENT APPLICATION NUMBER: US/10/003,759
 C--> 13 <141> CURRENT FILING DATE: 2001-10-23
 13 <150> PRIOR APPLICATION NUMBER: PCT/IS01/00012
 14 <151> PRIOR FILING DATE: 2001-06-15
 16 <150> PRIOR APPLICATION NUMBER: 09/594,884
 17 <151> PRIOR FILING DATE: 2000-06-15
 19 <160> NUMBER OF SEQ ID NOS: 10
 21 <170> SOFTWARE: FastSEQ for Windows Version 4.0
 23 <210> SEQ ID NO: 1
 24 <211> LENGTH: 1877
 25 <212> TYPE: DNA
 26 <213> ORGANISM: Rhodothermus marinus
 28 <220> FEATURE:
 29 <221> NAME/KEY: CDS
 30 <222> LOCATION: (710)..(1495)
 31 <223> OTHER INFORMATION:
 33 <400> SEQUENCE: 1
 34 gggcgagtgc aacgtcggc ggcgcctcgca gatcgAACGG ctcttccac gctgattcgc 60
 36 tgcgtgcgca cgcactggc gctggcgag ttcccttct acttcgtcca gattccccct 120
 38 tacgactacg ggccgaacgc ccgcacgcag cgtgtccggg aagcgcagtt ggcaccatg 180
 40 tgcgcgtgcc tcatacgggt atggctgtga ccaccatgt gggcgcacgac cacaacacta 240
 42 cctccggcgc gacaacgtga atggcgagc gcctgcctc tggcgctgg ccaacgacta 300
 44 cggttccgg gatgtggct actccggctc catctacgaa cgcatggAAC gtgaggatgg 360
 46 ccgcctgcgg ctctatttcc gctatgccgaa aggggactg gtcctgcgcc cggccgacgg 420
 48 cgccggcgag ttctgtatttgc cgggaccggc ccgtgtcttc cacccggcgc ggtacgggt 480
 50 cgaaggcgaa acgctggtg tttggagtcc ggcgtctcc gatccgcagg cggtcgccta 540
 52 tggctggtcc aatacgcgc acggccacgt gttcaaccgg gccggcctgc ccccttcgccc 600
 54 cttccgcacc gacgactggc cggaaaggcga ctgagcgacg caaccgggtgc ttgcattgcga 660
 56 caggggact tcgtaccttgc aagtgcctt ttttcattttca atggataaa atg aac gtc 718
 57 Met Asn Val
 58 1
 60 atg cgt gcg gta ctg gtc ctg agc ctg tta ttg ctg ttt gga tgc gac 766
 61 Met Arg Ala Val Leu Val Ser Leu Leu Leu Leu Phe Gly Cys Asp
 62 5 . 10 15
 64 tgg ctc ttt ccc gat ggc gac aac gga aag gaa ccg gag cct gag ccc 814
 65 Trp Leu Phe Pro Asp Gly Asp Asn Gly Lys Glu Pro Glu Pro Glu Pro
 66 20 25 30 35
 68 gag ccg acc gtc gag ctg tgc gga cgc tgg gac ggc cgc gat gtg gcc 862
 69 Glu Pro Thr Val Glu Leu Cys Gly Arg Trp Asp Ala Arg Asp Val Ala
 70 40 45 50

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/003,759

DATE: 12/14/2001
TIME: 09:52:58

Input Set : A:\2739.2001-001 seq.txt
Output Set: N:\CRF3\12142001\I003759.raw

72	ggg ggg cgc tac cgg gtg atc aac aac gta tgg ggc gcg gag acc gcc	910
73	Gly Gly Arg Tyr Arg Val Ile Asn Asn Val Trp Gly Ala Glu Thr Ala	
74	55 60 65	
76	cag tgc att gag gtc gga ctg gaa acg ggc aac ttc acg atc aca cgg	958
77	Gln Cys Ile Glu Val Gly Leu Glu Thr Gly Asn Phe Thr Ile Thr Arg	
78	70 75 80	
80	gcc gat cac gac aac ggc aac aac gtg gcc gcc tat ccg gcc atc tac	1006
81	Ala Asp His Asp Asn Gly Asn Asn Val Ala Ala Tyr Pro Ala Ile Tyr	
82	85 90 95	
84	tcc ggg tgc cac tgg ggc gcc tgc acg agc aat tcg gga ttg ccg cgg	1054
85	Phe Gly Cys His Trp Gly Ala Cys Thr Ser Asn Ser Gly Leu Pro Arg	
86	100 105 110 115	
88	cgc gtg cag gag ctg tcc gac gtg cgc acg agc tgg acg ctc acg ccg	1102
89	Arg Val Gln Glu Leu Ser Asp Val Arg Thr Ser Trp Thr Leu Thr Pro	
90	120 125 130	
92	atc acg acg ggc cgc tgg aat gcc gcc tac gac atc tgg ttc agt ccc	1150
93	Ile Thr Thr Gly Arg Trp Asn Ala Ala Tyr Asp Ile Trp Phe Ser Pro	
94	135 140 145	
96	gtc acg aat tcc ggc aac ggc tac agc ggc ggc gcc gag ctg atg atc	1198
97	Val Thr Asn Ser Gly Asn Gly Tyr Ser Gly Gly Ala Glu Leu Met Ile	
98	150 155 160	
100	tgg ctg aac tgg aac ggc ggc gtg atg ccg ggc agc cgc gtg gcc	1246
101	Trp Leu Asn Trp Asn Gly Gly Val Met Pro Gly Gly Ser Arg Val Ala	
102	165 170 175	
104	acc gtg gaa ctg gcc ggg gcc acc tgg gaa gtc tgg tat gcc gac tgg	1294
105	Thr Val Glu Leu Ala Gly Ala Thr Trp Glu Val Trp Tyr Ala Asp Trp	
106	180 185 190 195	
108	gac tgg aat tac atc gcc tac cgg cgc acg agc ccc acc acg tcg gtg	1342
109	Asp Trp Asn Tyr Ile Ala Tyr Arg Arg Thr Thr Pro Thr Thr Ser Val	
110	200 205 210	
112	agc gag ctg gac ctg aag gcc ttc atc gac gac gcg gtc gcc cgc ggc	1390
113	Ser Glu Leu Asp Leu Lys Ala Phe Ile Asp Asp Ala Val Ala Arg Gly	
114	215 220 225	
116	tac atc cgg ccg gag tgg tat ctg cat gcg gtg gag acg ggc ttc gaa	1438
117	Tyr Ile Arg Pro Glu Trp Tyr Leu His Ala Val Glu Thr Gly Phe Glu	
118	230 235 240	
120	ctc tgg gag ggc ggg gcc ggt ctg cga agc gcc gat ttt tcc gta acg	1486
121	Leu Trp Glu Gly Ala Gly Leu Arg Ser Ala Asp Phe Ser Val Thr	
122	245 250 255	
124	gtg cag tag cctgtcacac gggcaccaggc gttaggccaga gaagcacccg	1535
125	Val Gln	
126	260	
128	tcggggcgct tatgcgggctt ggcgtcgctt tgtgcctgac ttgttagtgc gctacggagg	1595
130	cgtcagccgg cgtggtgctt ttccctggag gcgcttcgtt tcgtgccggc cgagaactta	1655
132	tccgaaaagc aagggagaga cctggtaagc caggcgtttg cactgcgcac cagaagcagc	1715
134	ggtttgtagt aagccacgaa ggcgtccggc gaagtggcct caccgcgcac ggcacttcgc	1775
136	tgcgtccgt gcctgactac gagegcgttg gccgtgcttc aaacggcagg ctataaagct	1835
138	atacaggtcg aaatcgcgaa cttctccgac acttacaaac ac	1877
141	<210> SEQ ID NO: 2	

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/003,759

DATE: 12/14/2001
TIME: 09:52:58

Input Set : A:\2739.2001-001 seq.txt
Output Set: N:\CRF3\12142001\I003759.raw

142 <211> LENGTH: 261
 143 <212> TYPE: PRT
 144 <213> ORGANISM: Rhodothermus marinus
 146 <400> SEQUENCE: 2
 148 Met Asn Val Met Arg Ala Val Leu Val Leu Ser Leu Leu Leu Phe
 149 1 5 10 15
 152 Gly Cys Asp Trp Leu Phe Pro Asp Gly Asp Asn Gly Lys Glu Pro Glu
 153 20 25 30
 156 Pro Glu Pro Glu Pro Thr Val Glu Leu Cys Gly Arg Trp Asp Ala Arg
 157 35 40 45
 160 Asp Val Ala Gly Gly Arg Tyr Arg Val Ile Asn Asn Val Trp Gly Ala
 161 50 55 60
 164 Glu Thr Ala Gln Cys Ile Glu Val Gly Leu Glu Thr Gly Asn Phe Thr
 165 65 70 75 80
 168 Ile Thr Arg Ala Asp His Asp Asn Gly Asn Asn Val Ala Ala Tyr Pro
 169 85 90 95
 172 Ala Ile Tyr Phe Gly Cys His Trp Gly Ala Cys Thr Ser Asn Ser Gly
 173 100 105 110
 176 Leu Pro Arg Arg Val Gln Glu Leu Ser Asp Val Arg Thr Ser Trp Thr
 177 115 120 125
 180 Leu Thr Pro Ile Thr Thr Gly Arg Trp Asn Ala Ala Tyr Asp Ile Trp
 181 130 135 140
 184 Phe Ser Pro Val Thr Asn Ser Gly Asn Gly Tyr Ser Gly Gly Ala Glu
 185 145 150 155 160
 188 Leu Met Ile Trp Leu Asn Trp Asn Gly Gly Val Met Pro Gly Gly Ser
 189 165 170 175
 192 Arg Val Ala Thr Val Glu Leu Ala Gly Ala Thr Trp Glu Val Trp Tyr
 193 180 185 190
 196 Ala Asp Trp Asp Trp Asn Tyr Ile Ala Tyr Arg Arg Thr Thr Pro Thr
 197 195 200 205
 200 Thr Ser Val Ser Glu Leu Asp Leu Lys Ala Phe Ile Asp Asp Ala Val
 201 210 215 220
 204 Ala Arg Gly Tyr Ile Arg Pro Glu Trp Tyr Leu His Ala Val Glu Thr
 205 225 230 235 240
 208 Gly Phe Glu Leu Trp Glu Gly Gly Ala Gly Leu Arg Ser Ala Asp Phe
 209 245 250 255
 212 Ser Val Thr Val Gln
 213 260
 216 <210> SEQ ID NO: 3
 217 <211> LENGTH: 786
 218 <212> TYPE: DNA
 219 <213> ORGANISM: Rhodothermus marinus
 221 <400> SEQUENCE: 3
 222 atgaacgtca tgcgtcggt actggtcctg agcctgttat tgctgtttgg atgcgactgg 60
 224 ctctttcccg atggcaca ccggaaaggaa ccggagcctg agcccgagcc gaccgtcgag 120
 226 ctgtgcggac gctgggacgc gcgcgtatgtg gccggggggc gctaccgggt gatcaacaac 180
 228 gtatgggcgc cggagaccgc ccagtgcatt gaggtcggac tggaaacggg caacttcacg 240
 230 atcacacggg ccgatcacga caacggcaac aacgtggccg cctatccggc catctacttc 300
 232 gggtgccact ggggcgcctg cacgagcaat tcgggattgc cgccggcgctg gcaggagctg 360

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/003,759

DATE: 12/14/2001
TIME: 09:52:58

Input Set : A:\2739.2001-001 seq.txt
Output Set: N:\CRF3\12142001\I003759.raw

234	tccgacgtgc	gcacgagctg	gacgctcacg	ccgatcacga	cgggccgctg	aatgccgc	420
236	tacgacatct	gttcaagtcc	cgtcacgaat	tccgcaacg	gctacagcgg	cggcccgag	480
238	ctgtatgatct	ggctgaactg	gaacggcg	gtgtatgccg	gccccggc	cgtggccacc	540
240	gtggaaactgg	ccggggccac	ctgggaagtc	tggtatgccg	actgggactg	gaattacatc	600
242	gcctaccggc	gcacgacg	caccacgtcg	gtgagcgg	tggacctgaa	ggccttcatc	660
244	gacgacg	tcgcccgg	ctacatccgg	ccggagtgg	atctgcatgc	gtggagacg	720
246	ggcttcgaac	tctggagagg	cggggcccgt	ctgcgaagcg	ccgattttc	cgtaacgg	780
248	cagtag						786
251	<210>	SEQ ID NO:	4				
252	<211>	LENGTH:	1134				
253	<212>	TYPE:	DNA				
254	<213>	ORGANISM:	Pyrococcus furiosus				
256	<220>	FEATURE:					
257	<221>	NAME/KEY:	exon				
258	<222>.	LOCATION:	(82)..(1041)				
259	<223>	OTHER INFORMATION:					
261	<400>	SEQUENCE:	4				
262	aatgtcggg	gcagattacg	aagagagaag	agcttttt	tataccttat	attcaaatct	60
264	aacaaatttg	gaggatgtcc	a	atg	agc	aag	111
265				aaa	aag	ttc	Met Ser Lys Lys Lys Phe Val Ile Val Ser
266				1	5	10	
268	atc tta aca atc ctt tta	gta cag gca	ata tat ttt	gta gaa aag	tat		159
269	Ile Leu Thr Ile Leu Leu Val Gln Ala Ile Tyr	Phe Val Glu Lys Tyr					
270	15	20	25				
272	cat acc tct gag gac aag tca act tca aat acc	tca tct aca cca ccc					207
273	His Thr Ser Glu Asp Lys Ser Thr Ser Asn Thr Ser	Thr Pro Pro					
274	30	35	40				
276	caa aca aca ctt tcc act acc aag gtt ctc aag att	aga tac cct gat					255
277	Gln Thr Thr Leu Ser Thr Thr Lys Val Leu Lys Ile Arg Tyr Pro Asp						
278	45	50	55				
280	gac ggt gag tgg cca gga gct cct att gat aag gat	ggt gat ggg aac					303
281	Asp Gly Glu Trp Pro Gly Ala Pro Ile Asp Lys Asp Gly Asp Gly Asn						
282	60	65	70				
284	cca gaa ttc tac att gaa ata aac cta tgg aac att	ctt aat gct act					351
285	Pro Glu Phe Tyr Ile Glu Ile Asn Leu Trp Asn Ile Leu Asn Ala Thr						
286	75	80	85	90			
288	gga ttt gct gag atg acg tac aat tta acc agc ggc	gtc ctt cac tac					399
289	Gly Phe Ala Glu Met Thr Tyr Asn Leu Thr Ser Gly Val Leu His Tyr						
290	95	100	105				
292	gtc caa caa ctt gac aac att gtc ttg agg gat aga	agt aat tgg gtg					447
293	Val Gln Gln Leu Asp Asn Ile Val Leu Arg Asp Arg Ser Asn Trp Val						
294	110	115	120				
296	cat gga tac ccc gaa ata ttc tat gga aac aag cca	tgg aat gca aac					495
297	His Gly Tyr Pro Glu Ile Phe Tyr Gly Asn Lys Pro Trp Asn Ala Asn						
298	125	130	135				
300	tac gca act gat ggc cca ata cca tta ccc agt aaa	gtt tca aac cta					543
301	Tyr Ala Thr Asp Gly Pro Ile Pro Leu Pro Ser Lys Val Ser Asn Leu						
302	140	145	150				
304	aca gac ttc tat cta aca atc tcc tat aaa ctt gag	ccc aag aac gga					591

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/003,759

DATE: 12/14/2001
TIME: 09:52:58

Input Set : A:\2739.2001-001 seq.txt
Output Set: N:\CRF3\12142001\I003759.raw

305	Thr Asp Phe Tyr Leu Thr Ile Ser Tyr Lys Leu Glu Pro Lys Asn Gly		
306 155	160	165	170
308 ctg cca att aac ttc gca ata gaa tcc tgg tta acg aga gaa gct tgg		639	
309 Leu Pro Ile Asn Phe Ala Ile Glu Ser Trp Leu Thr Arg Glu Ala Trp			
310 175	180	185	
312 aga aca aca gga att aac agc gat gag caa gaa gta atg ata tgg att		687	
313 Arg Thr Thr Gly Ile Asn Ser Asp Glu Gln Glu Val Met Ile Trp Ile			
314 190	195	200	
316 tac tat gac gga tta caa ccg gct ggc tcc aaa gtt aag gag att gta		735	
317 Tyr Tyr Asp Gly Leu Gln Pro Ala Gly Ser Lys Val Lys Glu Ile Val			
318 205	210	215	
320 gtc cca ata ata gtt aac gga aca cca gta aat gct aca ttt gaa gta		783	
321 Val Pro Ile Ile Val Asn Gly Thr Pro Val Asn Ala Thr Phe Glu Val			
322 220	225	230	
324 tgg aag gca aac att ggt tgg gag tat gtt gca ttt aga ata aag acc		831	
325 Trp Lys Ala Asn Ile Gly Trp Glu Tyr Val Ala Phe Arg Ile Lys Thr			
326 235	240	245	250
328 cca atc aaa gag gga aca gtg aca att cca tac gga gca ttt ata agt		879	
329 Pro Ile Lys Glu Gly Thr Val Thr Ile Pro Tyr Gly Ala Phe Ile Ser			
330 255	260	265	
332 gtt gca gcc aac att tca agc tta cca aat tac aca gaa ctt tac tta		927	
333 Val Ala Ala Asn Ile Ser Ser Leu Pro Asn Tyr Thr Glu Leu Tyr Leu			
334 270	275	280	
336 gag gac gtg gag att gga act gag ttt gga acg cca agc act acc tcc		975	
337 Glu Asp Val Glu Ile Gly Thr Glu Phe Gly Thr Pro Ser Thr Thr Ser			
338 285	290	295	
340 gcc cac cta gag tgg tgg atc aca aac ata aca cta act cct cta gat		1023	
341 Ala His Leu Glu Trp Trp Ile Thr Asn Ile Thr Leu Thr Pro Leu Asp			
342 300	305	310	
344 aga cct ctt att tcc taa atttcgcaa cctggaaatt atcaagttta		1071	
345 Arg Pro Leu Ile Ser			
346 315			
348 agaaaaggta gagttgctaa agaattcaa gaaaattga aaagtaactt ttattgtat		1131	
350 ctc		1134	
353 <210> SEQ ID NO: 5			
354 <211> LENGTH: 319			
355 <212> TYPE: PRT			
356 <213> ORGANISM: Pyrococcus furiosus			
358 <220> FEATURE:			
359 <221> NAME/KEY: SIGNAL			
360 <222> LOCATION: (1)..(19)			
361 <223> OTHER INFORMATION:			
363 <220> FEATURE:			
364 <221> NAME/KEY: DOMAIN			
365 <222> LOCATION: (28)..(49)			
366 <223> OTHER INFORMATION: Linker Moiety			
368 <220> FEATURE:			
369 <221> NAME/KEY: DOMAIN			
370 <222> LOCATION: (50)..(319)			

VERIFICATION SUMMARY
PATENT APPLICATION: US/10/003,759

DATE: 12/14/2001
TIME: 09:52:59

Input Set : A:\2739.2001-001 seq.txt
Output Set: N:\CRF3\12142001\I003759.raw

L:13 M:270 C: Current Application Number differs, Replaced Current Application No
L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date